Cost Prediction of Tunnel Construction Based on Interpretative Structural Model and Stacked Sparse Autoencoder

Jing-Qun Zhou*, Qi-Ming Liu, Chang-Xi Ma, and Dong Li

Abstract—**Cost management plays a vital role in ensuring the successful execution of different engineering projects, with precise costing serving as the cornerstone of effective cost management strategies. Recently, the machine learning technique offers an accurate and efficient method for forecasting construction expenses, introducing a novel approach to cost accounting other than conventional calculation techniques. This paper provides an overview of the current research landscape in the realm of cost prediction utilizing machine learning, also addresses some new research focuses and limitations. By utilizing highway tunnel engineering as a case study, this study employs an Interpretative Structural Model (ISM) to analyze the primary factors influencing construction costs. Subsequently, a construction cost prediction model is developed, leveraging a Stacked Sparse Autoencoder (SSAE) network within a deep learning framework. Last, the proposed model is trained using real construction projects as samples. Results show that there are some good prediction outputs with a remarkably low mean absolute percentage error of 0.71%. Thereby, they verified the identifying precision of key influencing factors and the reliability of the cost prediction model.**

Index Terms—**construction cost, highway tunnel engineering, Interpretative Structural Model, prediction, Stacked Sparse Autoencoder**

I. Introduction

COST management covers the entirety of project construction, including planning, decision-making, and other construction stages. It not only impacts the interests of project stakeholders but also has significant implications for governmental bodies, industry associations, and the broader public. Within the planning and decision-making stages, the cost minimization goal plays a pivotal role in guiding investment decisions and selecting implementation strategies. Throughout the implementation phase, effective cost management facilitates the rational distribution and utilization of project resources, aids in the timely identification of potential risks and challenges, and enables the coordinated oversight of project schedule, quality, safety, and environmental considerations. Thereby it enhances construction progress. Cost management is essential for successful project execution and the attainment of project objectives, with precise cost accounting serving as the cornerstone of effective cost management.

Currently, engineering construction costs are calculated using customized software, with the basic calculation principle as follows. Cost indicators for labor, machinery, and material consumption are determined by selecting cost indices based on entered quantities. Then, the engineers are multiplying them by specified labor, machinery, and material prices to calculate direct costs. Based on the direct costs, other costs are calculated using rate standards selected according to the characteristics of the project, and the total construction cost is obtained by summarizing direct costs and other costs. The current cost calculation methods are cumbersome, time-consuming, and inefficient, with the calculation process being primarily guided by subjective human understanding, relying on manual data processing. The resulting calculations frequently exhibit substantial discrepancies from actual expenditures, while the static and inflexible nature of these methods fails to accommodate the dynamic demands of construction projects [1]. With the advent of the big data era and the rapid development of artificial intelligence, variable, and feature selection is conducted from engineering design and construction parameters and work environment to predict construction costs [2]. Based on machine learning and construction cost management theory, these new methods are providing a new approach to obtaining construction costs. This predictive method not only provides immediate and accurate results, promptly capturing project changes for dynamic cost monitoring but also enables reasonably accurate cost predictions before commencing design tasks. The preemptive nature of this approach is of significant importance in guiding design and construction activities and cost control.

Over the past three decades, substantial advancements have been achieved in cost prediction research leveraging big data and machine learning, which can be succinctly categorized into three stages. In the first stage, the prediction of construction cost components such as cost indicators and material prices was conducted. This was

Manuscript received November 24, 2023; revised August 14, 2024.

This work was supported by the Double-First Class Major Research Programmes GSSYLXM-04, Educational Department of Gansu Province.

Jing-Qun Zhou is a postgraduate student at the School of Traffic and Transportation, Lanzhou Jiaotong University, LanZhou 730070, China. (Corresponding author, phone: 6860199, e-mail: 18143773065@163.com).

Qi-Ming Liu is the department head of Gansu Road and Bridge Construction Group, LanZhou 730070, China. (e-mail: 1850398588@qq.com).

Chang-Xi Ma is a professor at the School of Traffic and Transportation, Lanzhou Jiaotong University, LanZhou 730070, China. (e-mail: machangxi@mail.lzjtu.cn).

Dong Li is a postgraduate student at the School of Traffic and Transportation, Lanzhou Jiaotong University, LanZhou 730070, China. (e-mail: 854268596@qq.com).

done using data sequence-based methods including time series analysis, exponential smoothing, regression analysis, and grey system theory, for instance. Wang et al. [3] employed the GM(1,1) model from grey system theory to forecast the comprehensive unit cost of residential construction in the area, using the existing residential buildings at Qufu Normal University as a case study for the upcoming decade. Zhang [4] utilized time series analysis to examine the traits of the unit cost sequence in affordable housing projects across six regions in Xinjiang, China and developed an ARIMA model for predictive purposes. In this stage, researchers also used decision trees, random forests, and other methods to select calculation parameters of traditional cost calculation methods in order to achieve more accurate cost calculations. For instance, An et al. [5] used the random forest algorithm to select important factors influencing the construction of power transmission line projects as calculation parameters for traditional cost calculation methods and demonstrated the superiority of this method through examples. In the second stage, cost prediction models were established using cost data and information from completed projects, utilizing principles of shallow algorithms such as Support Vector Regression, Traditional Hidden Markov Models, and Backpropagation Neural Networks (BPNN). Jiang et al. [6], for instance, optimized the initial threshold values of BPNN using particle swarm algorithm, and based on grey relational analysis, constructed an index system to establish the PSO-BPNN. Li [7] utilized Support Vector Machines to establish the mapping relationship between engineering costs and influencing factors, optimized the algorithm parameters of Support Vector Machines through genetic algorithms and developed a cost prediction model for comprehensive utility tunnel projects. Yang [8] devised a cost prediction model for residential building projects using BP Neural Networks, which simulated and forecasted the construction costs of 26 residential building projects in Hangzhou City. Li et al. [9] optimized the parameters of the Extreme Learning Machine model using the Bird Swarm Algorithm to develop the BSA-ELM model for predicting construction costs of building projects. Diana et al. [10] integrated both the Bromilow's time cost model and process-based data-driven model to propose an early stage construction cost prediction model, which has shown the highest level of prediction accuracy to date. In the third stage, researchers endeavored to develop innovative prediction models by applying diverse theoretical principles in order to discover new methods for better prediction results. For example, Swei [11] presents an approach for cost estimation that combines a maximum likelihood estimator for data transformations with least angle regression for dimensionality reduction to evaluate competing transportation investments. Shahrara [12] employed gene expression programming techniques to create a prediction model for automating the estimation of construction costs for water and sewer replacement projects. Many scholars have summarized the development of cost prediction based on machine learning. For instance, Duan [13] summarized the principles, advantages, disadvantages, and application requirements of commonly used cost prediction methods. Additionally, the author outlined the progress of the research on engineering cost prediction models. Finally, Duan suggested improving the universality of models and strengthening the establishment of comprehensive market cost information databases containing various construction elements. Elmousalami [14] summarised the existing construction cost prediction models and emphasized that these models are all considered as black box models, which makes their generalization difficult.

In summary, there are two main shortcomings in current research. Firstly, scholars have not fully utilized more advanced deep learning theories to build deep neural networks with better predictive performance [15] for forecasting engineering construction costs. Secondly, due to the fixed structural form, minimal construction interference established technical management methods, and strong predictability of costs in residential and small-scale installation projects, current cost prediction studies based on machine learning mainly focus on building and installation engineering.

Highway engineering construction is significantly influenced by the natural environment, characterized by high site mobility, stringent quality and safety control requirements, and construction costs constrained by the level of construction organization management. Predicting costs for highway engineering presents more significant challenges compared to building projects, particularly for highway mountain tunnel projects, which represent the most intricate aspect of highway engineering due to numerous uncontrollable construction factors, extensive organizational coordination challenges, and high technical demands [16]. To date, no scholar has investigated cost prediction specifically for highway mountain tunnel projects. This paper boldly selects the construction cost of highway mountain tunnel projects as the research focus for prediction, conducts ISM analysis on the chosen prediction features, and utilizes SSAE to forecast the construction cost of highway tunnel projects.

II. Methodology

A. Interpretive Structural Model

The construction environment in highway tunnel engineering is intricate and dynamic, influenced by various factors that impact costs. These factors interact and impose constraints on one another, leading to a complex and intricate relationship. In employing the empirical method to choose cost-influencing factors as input variables for prediction, the unclear relationships and constraints among these factors can lead to the selection of numerous factors with overlapping relationships and consistent effects. Consequently, this results in an excessive selection of influencing factors, heightened model complexity, and diminished prediction accuracy. Employing the ISM method to elucidate the logical structure relationships among influencing factors, recognize the hierarchical relationships of each influencing factor, and pinpoint the key factors governing and determining construction costs can streamline the prediction process and enhance prediction accuracy.

The ISM method was proposed by Warfield in 1976 [17] and has been used extensively in the analysis of complex systems in engineering. The ISM method is based on the principles of interaction, causality and transitivity. It transforms intricate and disorderly relationships among system elements into a precise multi-level hierarchical structural model, facilitating the analysis and disclosure of complex relationship structures to enhance comprehension of the system's key factors. The steps in the application of the ISM method are shown in Figure 1.

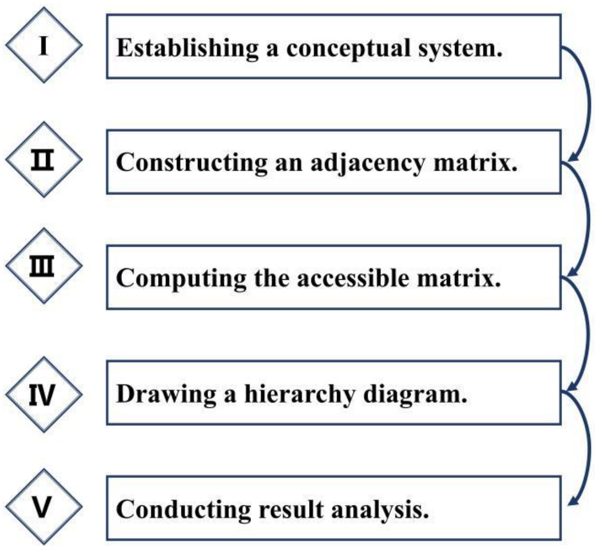

Fig. 1. Flowchart of the ISM method usage.

The conceptual system comprises system elements and the relationships that constrain them. The adjacency matrix, derived from the conceptual system and referred to as the original data matrix, is represented as a matrix $A=\left[\mathrm{a}_{ij}\right]_{n\times n}$. The matrix element a_{ij} is given by (1).

$$\mathrm{a}_{ij}=\begin{cases}1 & \mathrm{C_i}\text{ direcly influences or constrains }\mathrm{C_j}\\ 0 & \mathrm{C_i}\text{ doesn't direcly influences or constrains }\mathrm{C_j}\end{cases} \tag{1}$$

where $\mathrm{C_i}$ or $\mathrm{C_j}$ represent system elements.

The accessible matrix describes whether there is a path from one element to another element. Let the accessible matrix be denoted as the matrix $P=\left[p_{ij}\right]_{n\times n}$, which can be expressed as (2).

$$p_{ij}=\begin{cases}1 & \text{there is a path between }\mathrm{C_i}\text{ and }\mathrm{C_j}\\ 0 & \text{there is no path between }\mathrm{C_i}\text{ and }\mathrm{C_j}\end{cases} \tag{2}$$

where $\mathrm{C_i}$ or $\mathrm{C_j}$ is still an element of the system.

Common techniques for solving accessible matrices include the chain multiplication method, the power method, and the Warshall method. This paper employs the chain multiplication method to solve the accessible matrix. The identity matrix is denoted as I, and Boolean operations are applied to matrix $A+I$ until it meets the specified condition $(A+I)^{(k-1)}\neq(A+I)^{2k}=(A+I)^{(k+1)}=P$, then P matrix is designated as the accessible matrix. The connectivity of the accessible matrix is computed, leading to its division into multiple disjoint regions. The connectivity of the accessible matrix is computed, leading to its division into multiple disjoint regions. Subsequently, contraction operations are performed to acquire the strongly connected components, followed by removing of all forward edges to derive the skeleton matrix. The hierarchical division involves initially extracting the system elements that represent the final results and positioning them in the topmost layer. The iterative process persists until all elements have been extracted, completing the hierarchical division of the system and unveiling the hierarchical relationships among the elements. Lastly, the hierarchical relationship diagram of the system is drawn up. Upon acquiring the hierarchical relationship diagram, the findings can be analyzed in light of the specific problem scenario to enhance comprehension and facilitate more effective problem resolution.

B. *Stacked Sparse Autoencoder*

Deep learning is derived from studies on artificial neural networks. Currently, learning algorithms for neural networks are mainly aimed at lower-level network structures, referred to as shallow structure neural networks, which consist of a single input layer, one hidden layer, and one output layer. Conversely, networks with higher levels of non-linear operations are called deep-structure neural networks, such as a neural network with one input layer, three hidden layers, and one output layer. Deep learning obtains the main driving variables of input data through layer-by-layer learning algorithms, which can effectively represent complex high-dimensional functions, such as high-order functions, with high computational complexity, good generalization ability, and the ability to obtain multiple levels of extracted features for repeated use in similar but different tasks [18]. Typical deep learning models comprise convolutional neural networks, deep belief networks, and stacked autoencoder models. Convolutional neural networks are multi-layer perceptron neural networks that enforce constraints on network structure through feature extraction, mapping, and subsampling. They excel in various pattern recognition tasks, particularly image processing and computer vision. Deep trust networks consist of stacks of special forms of Boltzmann machines and are widely used in data dimensionality reduction, feature learning, collaborative filtering, and topic modeling. The cost-influencing factors in highway tunnel engineering are intricate, and acquiring engineering characteristics and cost data is challenging. This problem is label-free, with a small sample size, high dimensionality, and nonlinearity, where samples are independent and unrelated, rendering it unsuitable for prediction using convolutional neural networks and deep belief network models.

Stacked autoencoder networks comprise stacked structural units known as autoencoders. Traditional autoencoders mainly consist of encoding and decoding stages, and the structure is symmetric. If there are multiple hidden layers, the encoding and decoding stages contain an equal number of hidden layers. The encoding and decoding processes can be described as (3) and (4).

$$\text{encoding process: } h_1 = \sigma_e(w_1 x + b_1) \tag{3}$$
$$\text{decoding process: } y = \sigma_d(w_2 h_1 + b_2) \tag{4}$$

where w_1 and b_1 represent the encoding weights and biases, w_2 and b_2 denote the decoding weights and biases, σ_e signifies the commonly used non-linear transformations like Sigmoid, Tanh, and Relu, among others, and σ_d indicates the same non-linear or affine transformation employed in the encoding process [19]. Traditional autoencoders train by minimizing the reconstruction error $L(x,y)$. The loss function can be represented as $J(W,b)=\sum L(x,y)$. In addition to expressing the reconstruction error between x and y in the form of mean square error, as shown in (5), cross-entropy can also be an alternative, as illustrated in (6).

$$J(W,b)=\sum L(x,y)=\sum \| y-x \|_2^2 \tag{5}$$
$$J(W,b)=\sum L(x,y)=-\sum_{i=1}^{n}(x_i \log(y_i)+(1-x_i)\log(1-y_i)) \tag{6}$$

The above encoding and decoding process does not involve the label information of the input data, so traditional autoencoders are classified as unsupervised learning methods.

Based on traditional autoencoders, classical improved autoencoders include convolutional autoencoders, sparse autoencoders, shrinkage autoencoders, etc. Sparse autoencoders constrain the average activation value of hidden layer neurons' output based on traditional autoencoders, suppressing most of the outputs of hidden layer neurons, achieving a sparse effect in the network [20]. Using the KL divergence, it forces the average activation value of hidden layer neurons' outputs to be close to a given sparse value and adds it as a penalty term to the loss function. The penalty term can be defined as (7).

$$KL(\rho \| \hat{\rho}_j)=\sum_{j=1}^{s_2}\rho\log\frac{\rho}{\hat{\rho}_j}+(1-\rho)\log\frac{1-\rho}{1-\hat{\rho}_j} \tag{7}$$

where $\hat{\rho}_j$ represents the average activation value of hidden units for m samples, s_2 is the number of hidden neurons in the hidden layer, and the index j represents each neuron in the hidden layer in turn. The penalty term is the relative entropy between two Bernoulli Random Variables with ρ as the mean and $\hat{\rho}_j$ as the mean. Subsequently, the loss function of the sparse autoencoder can be expressed as (8).

$$J_{SAE}(W)=\sum L(x,y)+\beta\sum_{j=1}^{h}KL(\rho \| \hat{\rho}_j) \tag{8}$$

where β is employed to regulate the weight of the sparse penalty term, being capable of assuming values within the range of 0 to 1. The lower ρ is, the more influential the inhibitory effect, and usually takes a value close to 0.

The structure of the sparse autoencoder is depicted in Fig. 2. The SSAE is a deep neural network comprised of multiple sparse autoencoder structural units. It employs the greedy layer-by-layer idea to train each autoencoder model unsupervised, iteratively initializing the parameters of each layer, stacking the neural networks of each layer, and transforming into a deep supervised feedforward neural network, adjusting all parameters according to the supervision criteria. As the number of layers of sparse autoencoders increases, the learned feature representation of the original data becomes more abstract [21]. The working process of the SSAE is illustrated in Fig. 3.

Sparse autoencoders can learn the key features of source data, revealing distinctive characteristics and patterns, reducing data dimensions, implementing sparse data coding, and showcasing outstanding noise resistance and generalization performance. The input data processing in a stacked autoencoder proceeds sequentially, layer by layer, with each neural network layer extracting varying levels of features from the raw data. Features acquired in the higher neural network layers remain constant regardless of changing factors. Additionally, the network's learned function involves a higher level of non-linear operational combinations, enhancing its capacity to address complex problems and improve robustness. This study opts for a SSAE network model to forecast the construction cost of highway tunnel engineering.

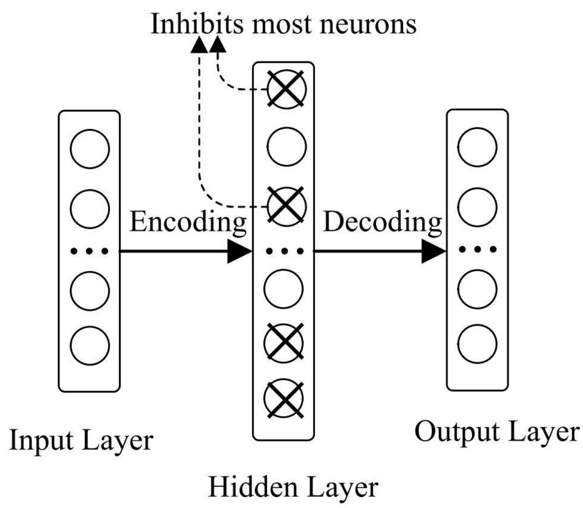

Fig. 2. Sparse autoencoder structure.

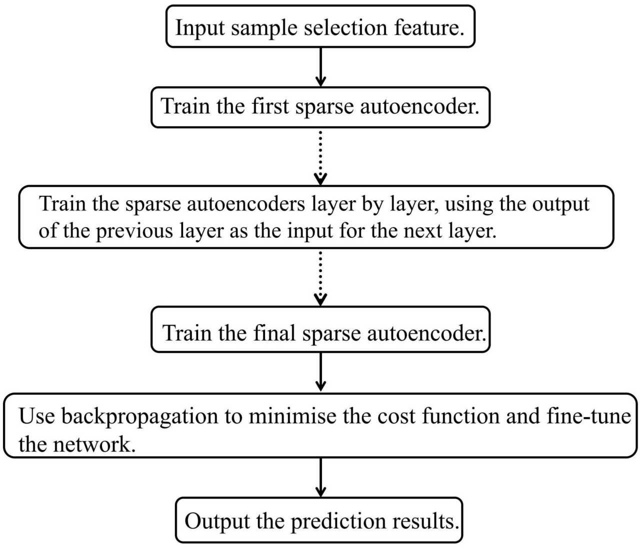

Fig. 3. The SSAE work process.

III. DATA

A. Identification of Influencing Factors

Employing methods such as expert interviews, historical data statistics, and field investigation methods, taking into account natural conditions like topography, geology, and hydrology, along with social and economic environments, tunnel construction conditions, design structure, and construction organization, comprehensively and completely identify the influencing factors on the construction cost of highway tunnel engineering from decision-making, surveying to completion. During the identification process, it is essential to adhere to the following principles: First, prioritize completeness by thoroughly identifying all factors influencing tunnel construction costs without omissions. Secondly, focus on low coupling to ensure minimal correlation between factors and a high degree of distinction. The identified factors should independently represent a class of influencing factors. Thirdly, emphasize conciseness by minimizing the number of factors while maintaining completeness. This can be achieved by eliminating weakly effective factors and combining those with similar effects. Fourthly, maintain objectivity by accurately defining influencing factors, assessing their impact on construction costs, and avoiding subjective speculation about their influence. Lastly, consider identifiability by ensuring that influencing factors are easily identifiable, and measurable, and that data collection and processing are straightforward. During the specific identification work, parameters such as pavement cross slope, pavement profile design, and tunnel plane design, although essential design parameters have a feeble impact on the construction cost of tunnels. Following the principle of conciseness, they are not considered influencing factors. For instance, while climate conditions can influence tunnel construction costs, they are dictated by the engineering geographical location. The geographical location has already been identified as a cost influencing factor, following the low-coupling principle, climate conditions are not identified as influencing factors. The paper concentrates on tunnel engineering costs. Furthermore, tunnel lighting, power supply, and monitoring fall under highway electromechanical and traffic safety engineering per international standards. Thus, they are not within the scope of this study.

According to the above principles and methods, the following 16 cost-influencing factors have been identified. Table I delineates each cost impact factor, providing explanations for their meanings and the scope of content they encompass. Table I is placed at the end of the content and before the reference part.

B. Selection of Key Influencing Factors

The ISM is used to analyse and select the key cost drivers. The process is as follows. Firstly, we have established the conceptual system. The identified tunnel construction cost influencing factors are considered as system elements, numbered sequentially from C1 to C16. Influence and constraint relationships between the system elements are determined to establish a conceptual system, based on the descriptions of each element in the previous section. It is essential to analyze and discuss the relationships between each element in detail. For example, based on daily experience, regional economic and social development constrain the level of participating construction units. Higher levels of social and economic development in a region correspond to increased levels of participating construction units. However, upon comprehensive comprehension of investment management and contracting systems, it is evident that construction units involved in highway tunnel projects do not exhibit regional characteristics. All construction units are equally able to partake in tunnel construction projects across different regions. In addition, they use advanced production materials from different regions at a single cost during the production and construction phases. Therefore, regional economic and social development does not directly influence the participation level of construction units. The ultimate conceptual system established is depicted in Table II. Table II is placed at the end of the content and before the reference part.

$$A = \begin{pmatrix}
0&1&1&1&1&1&1&0&0&0&0&1&0&1&0&1\\
0&0&1&1&1&1&1&1&0&1&1&0&0&0&0&1\\
0&0&0&0&0&0&1&1&1&1&1&0&0&0&0&1\\
0&0&0&0&0&0&1&1&1&1&1&0&1&0&1&1\\
0&0&0&0&0&0&1&1&1&1&1&0&0&0&0&1\\
0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&1\\
0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&1\\
0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&1\\
0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&1\\
0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&1\\
0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&1\\
0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0\\
0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&1\\
0&0&0&0&0&0&0&0&0&0&0&1&0&0&0&1\\
0&0&0&0&0&0&0&0&0&0&0&1&1&0&0&1\\
0&0&0&0&0&0&0&0&0&0&0&1&0&0&0&0
\end{pmatrix} \tag{9}$$

Second, based on the relationships between elements in the conceptual system, determine the value of (i=1,2,....,16, and j=1,2,....,16), and construct adjacency matrix A, which can be seen in (9).

Then, using the method of continued multiplication, we obtain the accessible matrix P, which can be seen in (10). Based on the accessible matrix, computational software is utilized for regional division, loop judgment, edge reduction operations, and hierarchical division to obtain the hierarchical relationships within the system and the relationships between various elements in the system, culminating in the illustration of the system's hierarchical relationships depicted in Fig. 4.

Finally, by analyzing the distribution of elements across different levels, with each lower level representing the cause of the upper level and the bottom level indicating the initial cause of the system, while the upper level reflects the outcome of the preceding level, the root cause is identified as the element at the bottom level. The system's hierarchical relationship diagram is employed to reconstruct the hierarchical relationship diagram among various cost influencing factors [23], as depicted in Fig. 5. Fig. 5 is placed at the end of the content and before the reference part.

$$P = \begin{pmatrix} 1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1 \\ 0&1&1&1&1&1&1&1&1&1&1&1&1&0&1&1 \\ 0&0&1&0&0&0&1&1&1&1&1&1&0&0&0&1 \\ 0&0&0&1&0&0&1&1&1&1&1&1&1&0&1&1 \\ 0&0&0&0&1&0&1&1&1&1&1&1&0&0&0&1 \\ 0&0&0&0&0&1&0&0&0&0&0&1&0&0&0&1 \\ 0&0&0&0&0&0&1&0&0&0&0&1&0&0&0&1 \\ 0&0&0&0&0&0&0&1&0&0&0&1&0&0&0&1 \\ 0&0&0&0&0&0&0&0&1&0&0&1&0&0&0&1 \\ 0&0&0&0&0&0&0&0&0&1&0&1&0&0&0&1 \\ 0&0&0&0&0&0&0&0&0&0&1&1&0&0&0&1 \\ 0&0&0&0&0&0&0&0&0&0&0&1&0&0&0&0 \\ 0&0&0&0&0&0&0&0&0&0&0&1&1&0&0&1 \\ 0&0&0&0&0&0&0&0&0&0&0&1&0&1&0&1 \\ 0&0&0&0&0&0&0&0&0&0&0&1&1&0&1&1 \\ 0&0&0&0&0&0&0&0&0&0&0&1&0&0&0&1 \end{pmatrix} \tag{10}$$

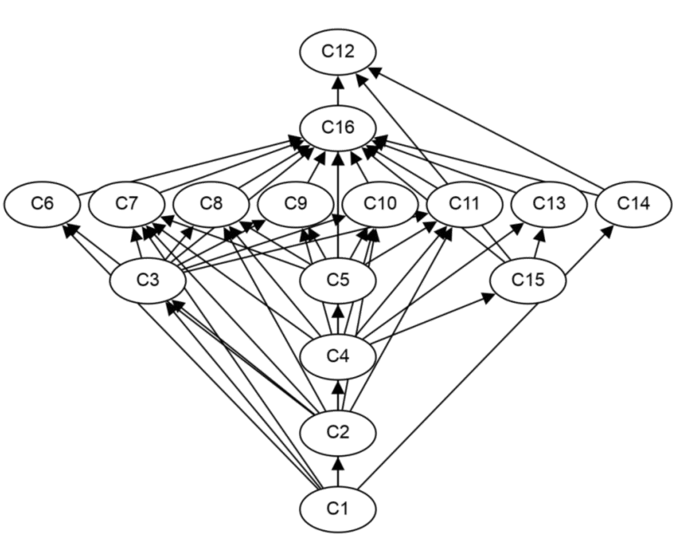

Fig. 4. Illustration of system hierarchical relationships.

Analyzing the interrelationships and constraints among various factors in the hierarchical diagram of cost influencing factors, geographical location, engineering geological conditions, and other natural conditions determine the tunnel construction, which in turn determines the three basic design parameters of tunnel length, tunnel type, and tunnel cross-section. The basic design parameters further determine the values of detailed design parameters such as tunnel portal structure, excavation support, tunnel trunk structure, and ventilating design. The top four layers of factors exert control over those beneath them, driving changes in lower layers and serving as the fundamental drivers of cost implications. Selecting the geographic position, engineering geological condition, tunnel length, tunnel type, and tunnel cross-section from the top 4 layers of influencing factors as input variables for the prediction model. Once the top 4 layers of influencing factors have been selected as input variables, there is no need to consider their lower-level influencing factors; otherwise, it will result in redundant model inputs. The level of participating units is a relatively independent influencing factor, which, apart from being affected by tunnel length, is unrelated to other factors. Its impact on cost needs to be considered separately, thus it is selected as an input variable for prediction.

C. Data Collection

Seventeen tunnel projects between 2020 and 2022 were collected and selected as the research sample, with construction starting in Gansu Province, China. These tunnels are situated in 17 distinct cities and counties within Gansu Province, with lengths varying from 214m to 6434m. They are invested in, designed, supervised, and constructed by owner units, design units, supervisory units, and construction units of different levels, ensuring a diverse and reliable dataset. Review of the design drawings, construction progress and cost information of the selected tunnel projects to gather data on key influencing factors. The original data are extracted and outlined in Table III. Table III is placed at the end of the content and before the reference part.

The chapter initially identifies the factors influencing tunnel construction costs using expert interviews, historical data statistics, and field investigations. Subsequently, employing the ISM, six key influencing factors are determined from a pool of 16 factors. These key factors include geographic location, engineering geological conditions, tunnel length, tunnel type, tunnel clearance section, and the level of participating units. Finally, sample tunnels are chosen to gather data on these key influencing factors.

IV. Experiment

A. Data Pre-processing

Six key influencing factors are categorized into quantitative and qualitative indicators. Ensure the qualitative index data is dimensionless, and normalize the quantitative index data.

Geographic position refers to the absolute geographical position of tunnel construction projects, determined by latitude and longitude. The geographical location of the project is divided by latitude and longitude into Central Gansu, Southern Gansu and Northern Gansu. The geographical location serves as a qualitative indicator with dimensionless units, with Central Gansu, Southern Gansu, and Northern Gansu corresponding to 1, 2, and 3 respectively.

The tunnel length refers to the average length of the left and right tunnels in the case of separate tunnels. If the tunnel is connected by a cut-and-cover tunnel, the length of these cut-and-cover tunnels is subtracted. The tunnel length is a quantitative indicator in meters.

Then, the influence of tunnel type and tunnel cross section on construction costs is considered, where they are categorizing the sampled tunnels into three groups. Based on this classification, detailed consideration is given to the parts of engineering geological conditions that have a significant impact on tunnel design, construction process, and working environment, such as the self-stability of surrounding rock, the hardness and integrity of rock mass, and the buried depth of the tunnel. By integrating the rock quality of the surrounding rock and relevant quality indicators, the tunnel sections are further categorized into 15 types, detailed in Table IV. Section engineering geological conditions in areas with significant deformations and active faults are unique and should be considered separately, classified into four types as shown in Table V. Based on a comprehensive consideration of the impact of engineering geological conditions, tunnel types, and tunnel cross sections on construction costs, the tunnel sections are divided into 19 types, which are numbered consecutively. Subsequently, the length of each section type in the sample tunnel is determined and utilized as input data for predictive analysis. The above-mentioned tunnel classification method is universally applicable, and any tunnel can be classified into sections using the process, as the basis for statistical prediction input data. Based on this classification criterion, statistical prediction input data is calculated. The length of each section type in the sample serves as a quantitative indicator, measured in meters.

Since the tunnel project as a sample is located in China, the classification standard of construction units in China is used as the basis for determining the grading standard of the participating units in this paper, as shown in Table VI, with excellent, good, fair, and poor grading scores of 1, 2, 3 and 4, respectively. Considering that the management level of the owner unit has an important influence on the level of the design, supervision and construction unit, the design and supervision unit guides and supervises the construction unit, and the construction unit's own construction technology and management level has a direct influence on the construction cost. We awarded the owner unit rating weight 0.3, design unit rating weight 0.2, supervision unit rating weight 0.2, construction unit rating weight 0.3. The overall rating of the engineering participating units is calculated by summing the weighted scores of the four components. The level of participating units is a quantitative indicator.

For quantitative factors, the dimensions of different quantitative factors are different, and the data differences within the same quantitative factor are too significant. If not processed, the data characteristics cannot be reflected. The dispersion standardization method is employed to normalizes each quantitative factor [24], limiting the input and output data from 0 to 1. The formula for the dispersion standardization transformation is as in (11).

$$\hat{x}_i = (x_i - x_{\min}) / (x_{\max} - x_{\min}) \tag{11}$$

where $\hat{x}_i$ is the normalized data; x_i is the currently collected data; $x_{\min}$ and $x_{\max}$ signify the minimum and maximum values respectively within the data for that category.

The non-dimensionalization and dispersion standardization methods mentioned above were implemented on the original sample data to prepare for the prediction work, with the outcomes detailed in Table VII. Table IV, table V, table VI and table VII are placed at the end of the content and before the reference part.

B. Modeling Prediction and Result Evaluation

As shown in Table VII, there are numerous qualitative indicators in the predicted input data represented by 0 and 1. However, data represented by 0 and 1 carry less information, and an excessive focus on these inputs by the model can lead to a decrease in prediction performance. Leveraging the principles and benefits of the SSAE detailed in Section 2, it can avoid interference from noisy data such as 0 and 1, enhance data feature extraction, and exemplify its proficiency in addressing complex high-dimensional non-linear problems. Consequently, employing the SSAE model for predicting the construction cost of highway tunnels is highly suitable.

The MATLAB platform is more concise than implementation platforms such as Python and C++, with less code, which is helpful for debugging. Therefore, in this paper we use MATLAB to establish an SSAE prediction model.

The SSAE prediction model utilized in this paper consists of an input layer, three autoencoder layers, a softmax layer, and an output layer. Fig. 6 illustrates the schematic diagram of this network model.

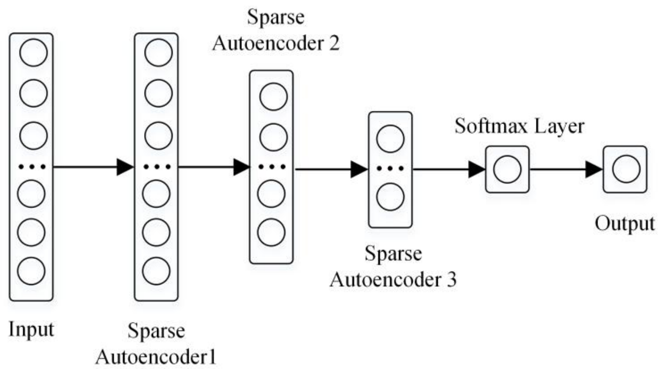

Fig. 6. The network structure of the prediction model.

The training samples, consisting of 13 training samples and 4 test samples, were randomly selected. When the number of neurons in the hidden layers of the three autoencoders is set to 300, 300, and 200, respectively, the maximum number of iterations is set to 500, 900, and 1000, respectively, the coefficients controlling the impact of sparse regularization in the cost function are set to 2, 4, and 3 respectively, and the transfer functions of the encoder and decoder are both set to the Logistic Sigmoid Function. The sparsity proportion is set to 0.05, and the L2 weight regularization coefficient in the cost function is set to 0.001. The model achieves the best training effect. After analysis, the three autoencoders reach their best training performance at the 500th, 92nd, and 95th iterations, respectively, with the corresponding performance values being 0.12502, 0.0018623, and 0.0000000051616. The calculated gradients are 0.000111, 0.00186, and 0.00000088, respectively. The training process of the three autoencoders is depicted in Figs. 7, 8, and 9. The final training results of the model are presented in Fig. 10, while the testing outcomes are depicted in Fig. 11. Through the analysis of Figs. 10 and 11, it is evident that there is a high degree of fit between the real and predicted values of the training and testing samples. The model exhibits outstanding performance, effectively capturing the relationships and trends in the data. It is rarely affected by noise interference and demonstrates strong generalization ability.

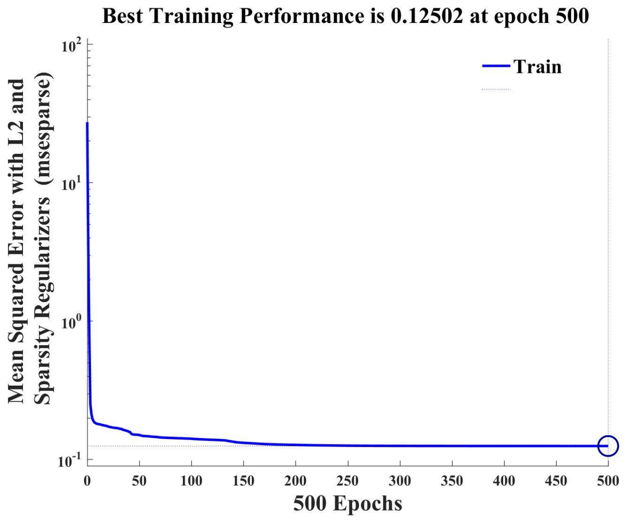

Fig. 7. Autoencoder1 training process.

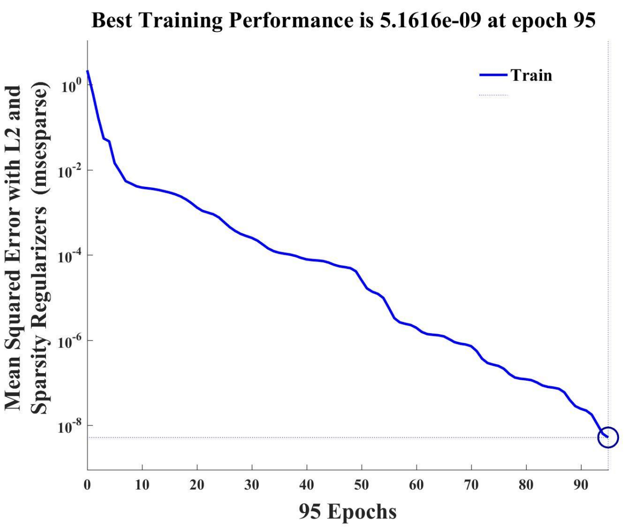

Fig. 8. Autoencoder2 training process.

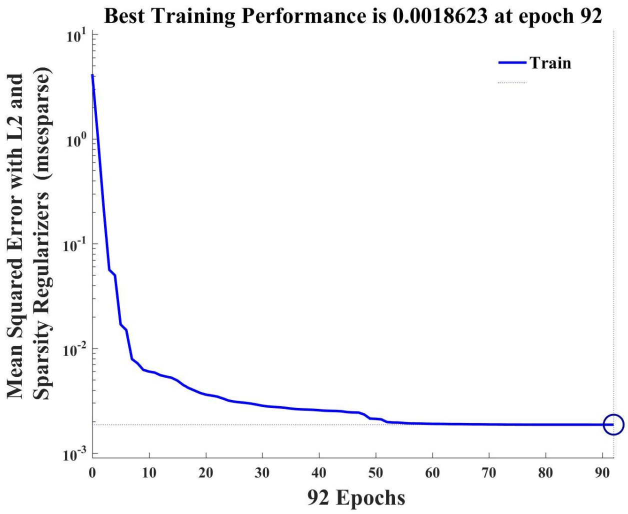

Fig. 9. Autoencoder3 training process.

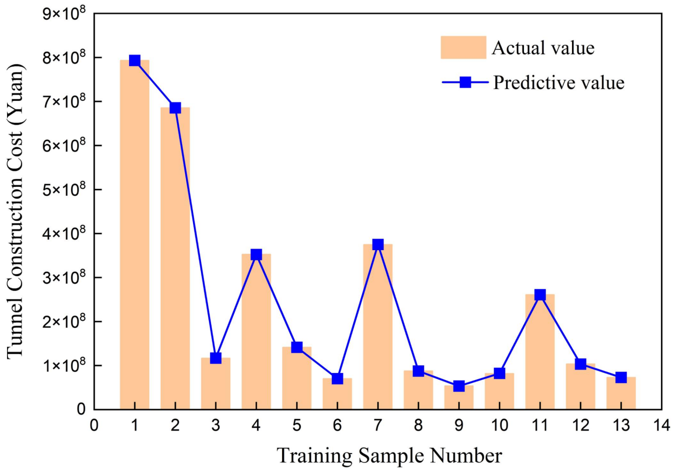

Fig. 10. The training effect of the SSAE model.

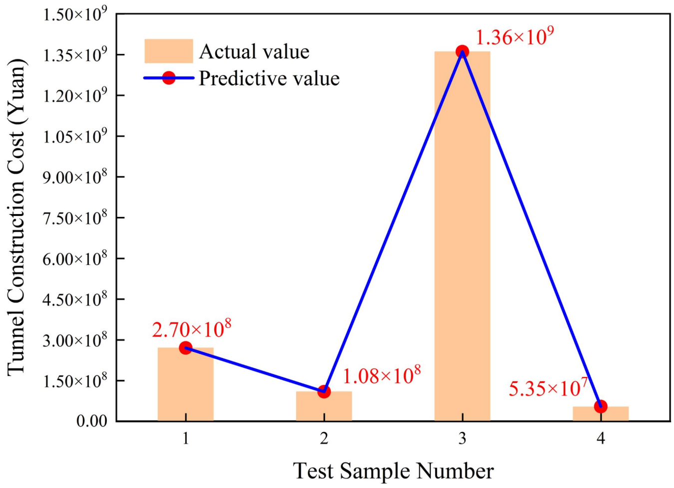

Fig. 11. The testing effect of the SSAE model.

To compare the predictive performance of deep and shallow neural network structures, we selected the best-fitting BP Neural Network (BPNN) model from the shallow structure models. Simulation predictions were conducted using Matlab with the same testing and training samples. The BP network structure is a two-layer feed-forward network with Sigmoid hidden neurons and linear output neurons. The number of hidden neurons was set to 102. Since the input matrix of the model contains a large number of zero elements and involves solving sparse linear equations, the conjugate gradient method is the most excellent iterative solution method for sparse linear equation groups. The training algorithm is selected as the Scaled Conjugate Gradient. The finalized testing outcomes are illustrated in Figure 12.

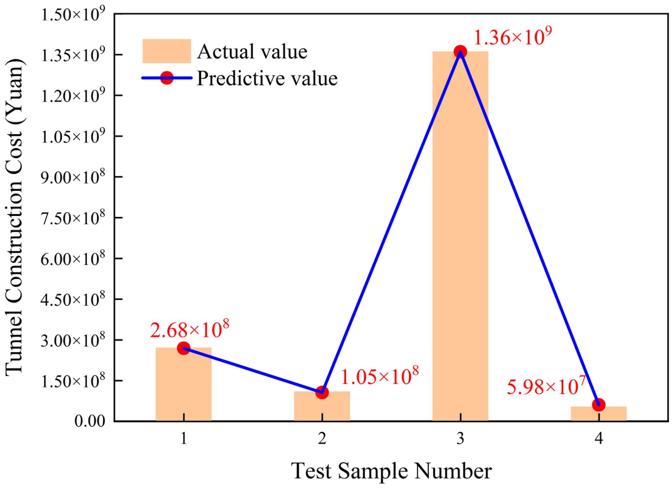

Fig. 12. The testing effect of the BPNN Model

Due to the large order of magnitude of the sample data, the prediction error is small relative to its volume, and the prediction error cannot be directly observed from Figures 11 and 12. To better compare the prediction performance of the models, the linear fit diagrams of the SSAE and BPNN models are plotted with the actual values as the horizontal axis and the predicted values as the vertical axis, as shown in Figures 13 and 14, respectively. It is clear from the linear fit diagrams that the Pearson's r of the SSAE model is 1 and the Pearson's r of the BPNN model is 0.9997, and the SSAE model has a better fit.

To precisely quantify and compare the prediction accuracy and specific disparities between the two prediction methods, the prediction outcomes of both approaches were de-normalized. Subsequently, the absolute error, relative error, root mean square error (RMSE), and mean absolute percentage error (MAPE) of the two models were computed and detailed in Table Ⅷ. Table Ⅷ is placed at the end of the content and before the reference part. The mean absolute percentage error of the prediction results from the BPNN model is below 5%, and the root mean square error is under 5 million yuan. Compared with its fitting effect on other cost datasets [8], the RMSE and MAPE of this dataset are significantly reduced, which verifies that selecting key influencing factors for cost is appropriate and plays an essential positive role in improving prediction accuracy. The SSAE model's predicted absolute error falls within the range of [-295923.32, 1297151.97], with a relative error between [-0.0075, 2.4827], an RMSE of 675842.24 ten thousand yuan, and a MAPE of 0.71%. The BPNN model exhibits an absolute error within the range of [-3199795.17, 7539099.72], a relative error between [-0.3886, 14.4293], a MAPE of 3.7349%, and an RMSE of 4131027.35 ten thousand yuan. The SSAE has a more stable prediction effect and higher prediction accuracy. The SSAE model has three sets of test samples with a predicted relative error controlled within 0.30%, and the most minor relative error of the sample is only 0.0075%. For a sample with an actual cost of 1.36 billion yuan, the error margin is merely 10.3 ten thousand yuan, which is nearly insignificant in actual engineering project construction. Please refer to Figure 15 and Figure 16 for a comparative visualization of the predicted relative and absolute errors from the two models. The error curve representing the prediction effect of the BPNN model has large ups and downs and a wide range of fluctuations, while the error curve representing the prediction effect of the SSAE model has small fluctuations, with the maximum relative error not exceeding 2.5% and the maximum absolute error not exceeding 129.72 ten thousand yuan. The comparison results show that the SSAE model is more stable and reliable.

Based on the above analysis, the tunnel construction cost prediction model established in this paper using SSAEs yielded an RMSE of 675,842.24 million RMB and a MAPE of 0.71%, with a minimum absolute relative error value of only 0.0075%. The prediction accuracy is high, reliable, and stable, surpassing models established on conventional shallow networks such as the BPNN.

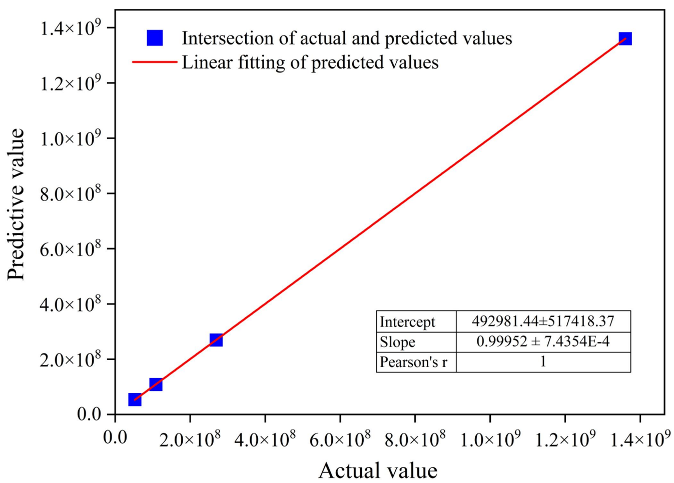

Fig. 13. Linear fitting of SSAE model predictions

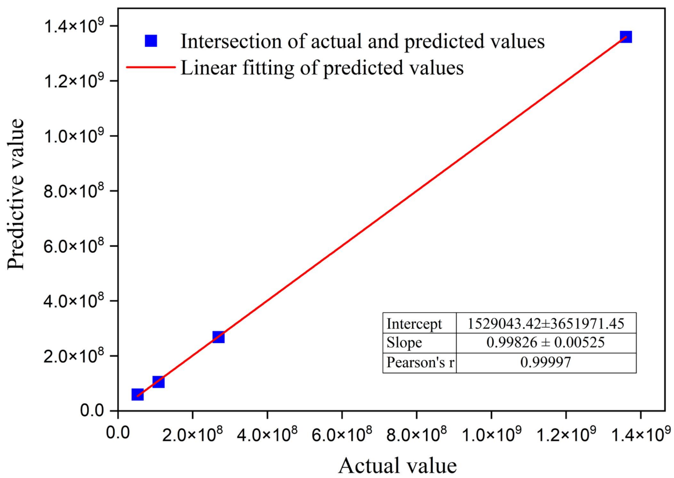

Fig. 14. Linear fitting of BPNN model predictions

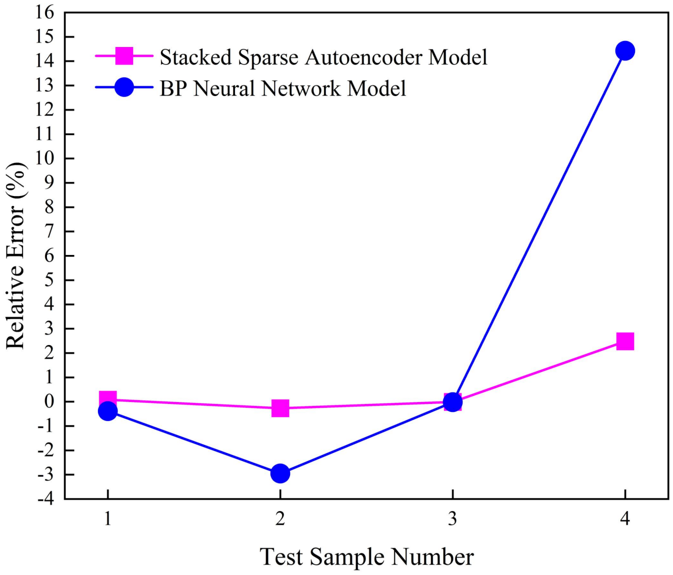

Fig. 15. Comparison of the predicted relative errors of the two models.

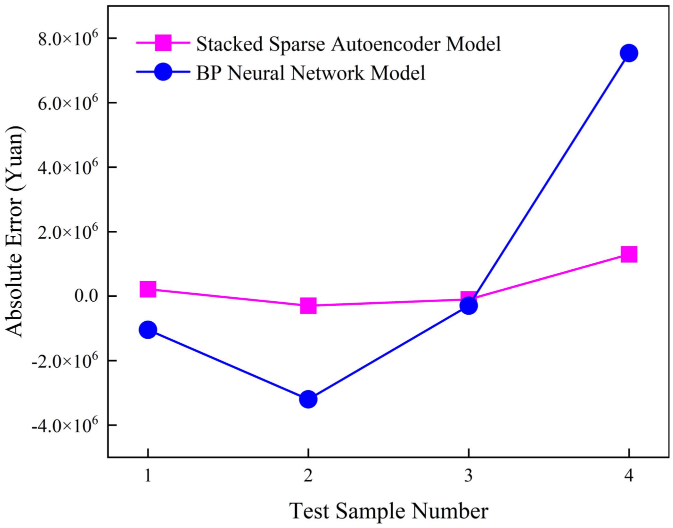

Fig. 16. Comparison of the predicted absolute errors of the two models.

V. Conclusion

This paper explores a new approach to determining construction costs for highway tunnel engineering using ISM methods and SSAE models. Firstly, the factors influencing tunnel construction costs were comprehensively identified, and using the ISM method, the influencing factors that play a controlling and determining role were identified and used as input variables for the predictive model. Secondly, an analysis was carried out to compare the advantages and disadvantages of shallow and deep neural network structures, to assess the suitability of common deep learning models, and ultimately to select the SSAE model for predicting tunnel construction costs. Finally, using real engineering projects as research samples, we extracted and processed the prediction input data, trained the prediction network model, derived the prediction results, and compared the prediction effect with the model constructed based on the principle of the BP neural network.

The main research findings are as follows:

1) The predictive model input variables selected by the interpretive structural model method have a strong correlation with construction costs, and they are comprehensive, concise, and have low coupling, thereby simplifying the predictive structural model and enhancing prediction accuracy.
2) The model, based on the SSAE, with a MAPE of 0.71%, shows that prediction accuracy is high, stable, and reliable.

TABLE I
INTRODUCTION TO THE SELECTED IMPACT FACTORS

Serial Number	Name of Cost Impact Factor	Definition of Impact Factor
C1	Geographical location.	The absolute geographical location where the tunnel project is located to determine natural conditions such as topography and climate.
C2	Engineering geological conditions.	Encompass three components. The first pertains to topography, lithology, geological structure, bad geology and special rocks and soils. The second component includes hydrogeological conditions such as hydrological conditions, water corrosivity evaluation, and predicted water inflow. The third component relates to the earthquake and neotectonic activity, primarily focusing on fault activity and seismic activity.
C3	Tunnel types.	Categorized based on the tunnel's cross-sectional layout, it includes single tunnel with bidirectional traffic, separated tunnel with unidirectional traffic, small interval tunnel, double arch tunnel, and bifurcated tunnel.
C4	Tunnel length.	The distance between the end wall and the wall surface of the tunnel entrance and exit doors.
C5	Tunnel clearance section.	Denotes the area and shape of the section on the inside of the tunnel lining.
C6	Tunnel portal structure.	Refers to the tunnel portal type, structure, and engineering. Tunnel portal types include end wall, wing wall, column, among others. Tunnel portal engineering encompasses the side slope, retaining wall, and drainage engineering of the portal.
C7	Excavation support and tunnel structure [22].	Pertains to tunnel excavation blasting, advance support, initial support, primary lining, secondary lining, inverted arch lining, and inverted arch backfill.
C8	Ventilation design.	Involves calculating tunnel air demand and developing a ventilation scheme, which may encompass natural ventilation, fans, shafts, inclined shafts, etc.
C9	Pavement design.	Encompasses tunnel pavement types, pavement structure design, and related aspects.

CONTINUATION OF TABLE I
INTRODUCTION TO THE SELECTED IMPACT FACTORS

Serial Number	Name of Cost Impact Factor	Definition of Impact Factor
C10	Auxiliary structures.	Refers to structures constructed for operation management, maintenance, water supply, drainage, ventilation, and safety. It mainly refers to the length and design structure of the open-cut tunnel, the traversing tunnel of the car, the traversing tunnel of the pedestrian and the emergency parking strip.
C11	Dynamic design and information construction.	Primarily focuses on tunnel site monitoring measurements and advanced geological predictions for the tunnel.
C12	Material prices.	Encompasses the prices of major outsourced materials like steel bars and cement, comprehensive outsourced materials such as iron wire and coatings, and local materials like sand and gravel.
C13	Waste soil and concrete transportation distance.	Refers to the distance from the tunnel entrance and body excavation site to the waste disposal site or spoil ground. The distance from the concrete mixing station to the construction site for tunnel invert, primary lining, secondary lining, and other concrete engineering tasks.
C14	Local economic and social development.	Encompasses the economic development level, local policy requirements, and social environment of the construction site area.
C15	Level of participating units.	Indicates the design, construction, organization, and management proficiency of the owner unit, design unit, supervision unit, and construction unit.
C16	Construction period.	The project completion deadline, which is the duration from the commencement of construction to the project's completion.

TABLE II
CONCEPTUAL SYSTEM

Serial Number	Systematic Element	Elements of Direct Impact or Direct Constraint	Serial Number	Systematic Element	Elements of Direct Impact or Direct Constraint
C1	Geographic position	Engineering geological condition, Tunnel type, Tunnel length, Tunnel cross section, Tunnel portal structure, Excavation support and tunnel trunk structure, Material price, Local economic and social development, Construction period	C9	Pavement design	Construction period
C2	Engineering geological condition	Tunnel type, Tunnel length, Tunnel cross section, Tunnel portal structure, Excavation support and tunnel trunk structure, Ventilating design, Ancillary structures, Dynamic design and informative construction, Construction period	C10	Auxiliary structures	Construction period
C3	Tunnel type	Excavation support and tunnel trunk structure, Ventilating design, Pavement design, Ancillary structures, Dynamic design and informative construction, Construction period	C11	Dynamic design and informative construction	Construction period
C4	Tunnel length	Excavation support and tunnel trunk structure, Ventilating design, Pavement design, Ancillary structures, Dynamic design and informative construction, Waste soil and concrete transportation distance, Level of participating units, Construction period	C12	Material price	
C5	Tunnel cross section	Excavation support and tunnel trunk structure, Ventilating design, Pavement design, Ancillary structures, Dynamic design and informative construction, Construction period	C13	Waste soil and concrete transportation distance	Construction period
C6	Tunnel portal structure	Construction period	C14	Local economic and social development	Material price, Construction period
C7	Excavation support and tunnel trunk structure	Construction period	C15	Level of participating units	Material price, Waste soil and concrete, Transportation distance, Construction period
C8	Ventilating design	Construction period	C16	Construction period	Material price

TABLE III
ORIGINAL DATA

Tunnel Number	Geographic Position		Total Tunnel Length (m)	Section Length (m)								Owner Level	Design Level	Supervision Level	Construction Level	Construction Cost (ten thousand yuan)
	Latitude	Longitude		F5a	F5b	F5c	F5d	....	XK5a	XK5b	XK4a					
1	100.46	38.94	5254.42	23	394	1195	0	....	0	0	0	1	1	1	1	79316.18
2	105.74	34.58	3753.50	0	79	1146	0	....	0	0	0	2	1	1	1	68565.31

CONTINUATION OF TABLE III
ORIGINAL DATA

Tunnel Number	Geographic Position		Total Tunnel Length (m)	Section Length (m)								Owner Level	Design Level	Supervision Level	Construction Level	Construction Cost (ten thousand yuan)
	Latitude	Longitude		F5a	F5b	F5c	F5d	....	XK5a	XK5b	XK4a					
3	102.09	38.44	410.5	0	0	0	0	....	311	490	0	2	1	1	2	11675.78
4	104.94	33.43	2402.50	5	251	251	0	....	0	0	0	2	2	1	1	35269.15
5	96.00	40.42	440.00	0	0	0	0	....	342.6	520	0	1	2	1	1	10842.35
6	103.70	36.44	693.50	15	0	0	0	....	0	0	0	3	2	2	2	14143.34
7	104.51	36.85	373.50	0	0	0	0	....	179	0	567	3	2	3	3	7022.98
8	103.22	35.60	2223.50	0	0	2995	0	....	0	0	0	2	2	1	1	37486.37
9	105.15	34.22	234.00	0	0	0	0	....	187	116	140	2	2	2	2	5401.88
10	104.88	34.69	346.00	0	0	0	0	....	0	409	0	2	1	2	2	8184.48
11	103.03	37.90	2326.50	0	331	1112	0	....	0	0	0	2	1	1	1	26933.91
12	104.63	35.06	2231.50	16	72	448	169	....	0	0	0	2	2	1	1	26088.81
13	99.84	39.14	745.00	10	128	113	0	....	0	0	0	1	1	1	1	10350.94
14	95.77	40.51	214.00	0	0	0	0	....	252	160	0	2	1	1	2	5224.85
15	105.08	35.72	550.50	0	0	0	0	....	0	0	0	3	2	2	3	7295.99
16	98.92	39.97	545.00	24	505	0	0	....	130	0	0	2	1	2	1	8782.14
17	104.70	32.95	6434.00	4	267	859	666	....	0	0	0	1	1	1	1	136014.55

TABLE IV
CLASSIFICATION OF TUNNEL SECTIONS BASED ON COST-INFLUENCING FACTORS

Classification by Type of Tunnel	Classification by Engineering Geological Conditions	Numbering
Standard section separation tunnel	Grade V surrounding rock semi-bright and semi-dark section and Grade V surrounding rock portal shallow buried bias section	F5a
	Grade V surrounding rock portal shallow buried reinforced section	F5b
	Grade V surrounding rock tunnel trunk normal section	F5c
	Grade V surrounding rock fault fracture zone section	F5d
	Grade IV surrounding rock shallow buried section	F4a
	Grade IV surrounding rock tunnel trunk normal section	F4b
	Grade IV surrounding rock deep buried section	F4c
	Grade III surrounding rock tunnel trunk normal section	F3a
Small interval tunnel with standard sections	Grade V surrounding rock portal shallow buried bias section and Grade V surrounding rock shallow buried reinforced section of 10-20 m interval	X5a
	Grade V surrounding rock tunnel trunk normal section	X5b
	Grade V surrounding rock portal shallowly buried reinforced section of 6-10m interval	XX5a
	Grade IV surrounding rock tunnel trunk normal section	X4a
Small interval tunnel of the same width as the subgrade	Grade V surrounding rock semi-bright and semi-dark section and Grade V surrounding rock portal shallow buried reinforced section	XK5a
	Grade V surrounding rock tunnel trunk normal section	XK5b
	Grade IV surrounding rock tunnel trunk normal section	XK4a

TABLE V
CLASSIFICATION OF TUNNEL SECTIONS BASED ON COST-INFLUENCING FACTORS

Classification by Geological Structure	Classification based on Residual Engineering Geological Conditions	Numbering
Large deformation and active fault sections	Grade V surrounding rock strong large deformation section	BX5a
	Grade V surrounding rock medium large deformation section	BX5b
	Grade IV surrounding rock slightly large deformation section	BX4a
	Grade V surrounding rock passes through mobile moving fault section	BX5c

TABLE VI
GRADING STANDARD FOR PARTICIPATING UNITS

Name of Participating Units	Excellent	Good	Fair	Poor
Owner's unit	Government unit	Central enterprises	State-owned enterprises	
Design unit	Grade A Design Qualification for Highway Industry	Grade A Professional Design Qualification for Highway Tunnel	Grade B Professional Design Qualification for Highway Tunnel	Grade C Professional Design Qualification for Highway Tunnel
Supervisory unit	Grade A Comprehensive Qualification	Grade A Professional Qualification	Grade B Professional Qualification	
Construction unit	Special Grade General Contracting Qualification for Construction Projects	Grade A General Contracting Qualification for Construction Projects	Grade A Professional Contracting Qualification for Tunnel Engineering	Grade B Professional Contracting Qualification for Tunnel Engineering

TABLE VII
PREPROCESSED DATA

Tunnel Number	Total Tunnel Length (m)	Position	Level of Participating Units	F5a	F5b	F5c	F5d	...	XK5a	XK5b	XK4a	Construction Cost (ten thousand yuan)
1	0.810	3.000	0.000	0.958	0.780	0.399	0.000	...	0.000	0.000	0.000	79316.18
2	0.569	2.000	0.120	0.000	0.156	0.383	0.000	...	0.000	0.000	0.000	68565.31
3	0.032	3.000	0.240	0.000	0.000	0.000	0.000	...	0.908	0.942	0.000	11675.78
4	0.352	2.000	0.200	0.208	0.497	0.084	0.000	...	0.000	0.000	0.000	35269.15
5	0.036	3.000	0.080	0.000	0.000	0.000	0.000	...	1.000	1.000	0.000	10842.35
6	0.077	1.000	0.600	0.625	0.000	0.000	0.000	...	0.000	0.000	0.000	14143.34
7	0.026	1.000	1.000	0.000	0.000	0.000	0.000	...	0.522	0.000	1.000	7022.98
8	0.323	1.000	0.200	0.000	0.000	1.000	0.000	...	0.000	0.000	0.000	37486.37
9	0.003	2.000	0.520	0.000	0.000	0.000	0.000	...	0.546	0.223	0.247	5401.88
10	0.021	2.000	0.440	0.000	0.000	0.000	0.000	...	0.000	0.787	0.000	8184.48
11	0.340	3.000	0.120	0.000	0.655	0.371	0.000	...	0.000	0.000	0.000	26933.91
12	0.324	1.000	0.200	0.667	0.143	0.150	0.254	...	0.000	0.000	0.000	26088.81
13	0.085	3.000	0.000	0.417	0.253	0.038	0.000	...	0.000	0.000	0.000	10350.94
14	0.000	3.000	0.240	0.000	0.000	0.000	0.000	...	0.736	0.308	0.000	5224.85
15	0.054	1.000	0.720	0.000	0.000	0.000	0.000	...	0.000	0.000	0.000	7295.99
16	0.053	3.000	0.200	1.000	1.000	0.000	0.000	...	0.379	0.000	0.000	8782.14
17	1.000	2.000	0.000	0.167	0.529	0.287	1.000	...	0.000	0.000	0.000	136014.55

TABLE VIII
COMPARISON OF PREDICTION RESULTS FROM DIFFERENT MODELS

Test Sample Number	Actual Construction Costs (yuan)	Projected Construction Costs (yuan)		Absolute Error (yuan)		Relative Error (%)	
		SSAE	BPNN	SSAE	BPNN	SSAE	BPNN
13	269339133.99	269554464.95	268292414.80	215330.96	-1046719.20	0.0799	-0.3886
5	108423515.00	108127591.69	105223719.83	-295923.32	-3199795.17	-0.2729	-2.9512
10	1360145463.01	1360042946.27	1359846777.34	-102516.75	-298685.67	-0.0075	-0.0220
16	52248518.00	53545669.97	59787617.72	1297151.97	7539099.72	2.4827	14.4293
RMSE				675842.24	4131027.35		
MAPE(%)						0.7108	4.4478

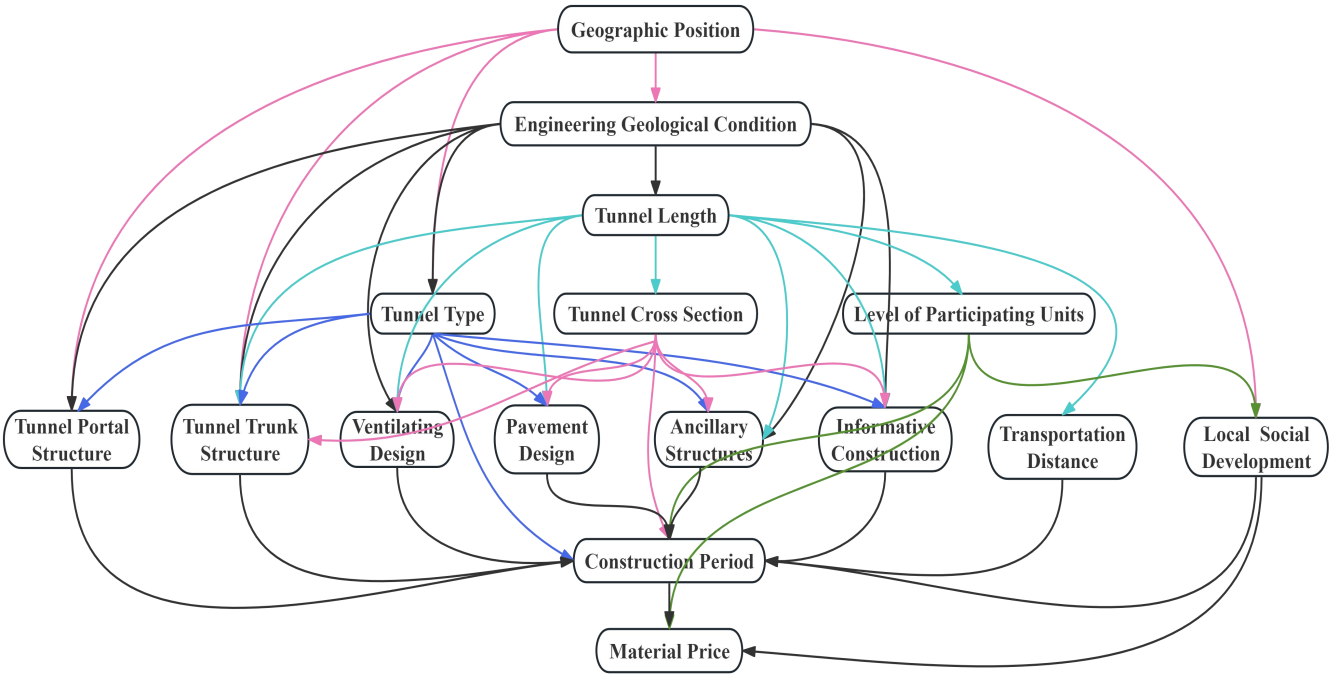

Fig. 5. Illustration of cost influencing factors hierarchical relationships.

REFERENCES

[1] Ming-Kang Li, "Comparative study on construction cost estimate algorithms based on machine learning," M.S. thesis, SCUT, GuangZhou, China, 2022.
[2] Guyon, Isabelle, and Andre Elisseeff. "An introduction to variable and feature selection," Journal of machine learning research, vol. 03, no. 3, pp.1157-1182, 2003.
[3] Min Wang, Yu-Qiang Wang, and Dong-Hui Wang. "Application of GM(1,1) model in the prediction of unilateral cost of residential project," Journal of Shandong Architectural Engineering Institute, vol. 04, no. 1, pp.39-43, 2000.
[4] Shan-Shan Zhang, "Research on the forecast of per square cost for security housing projects in XinJiang based on time series analysis," M.S. thesis, XJU, XinJiang, China, 2021.
[5] Lei An, Jie Zhang, and Xia Qi, "Research on an estimation of project cost of transmission and substation lines based on random forest," Control Engineering, vol. 23, no. 11, pp.1841-1844, 2016.
[6] Hong-Yan Jiang and Yu-Qing Bai, "Estimation of high-rise residential cost based on PSO-BP neural network with gray correlation," Journal of Engineering Management, vol. 33, no. 1, pp.29-33, 2019.
[7] Qian Li and You Zhang, "Research on the estimation method of civil engineering cost of comprehensive pipeline corridor based on genetic support vector machine," Tunnel Construction, vol. 38, no. 2, pp.171-175, 2018.
[8] Jin-Yue Yang, "Research on construction project cost prediction based on BP neural network," M.S. thesis, ZJU, ZheJiang, China, 2015.
[9] Wan-Qing Li, Jia-Qi Chen, and Wen-Qing Meng, "Research on construction cost prediction of building project based on BSA-ELM model," Practice and Understanding of Mathematics, vol. 49, no. 23, pp.10-17, 2019.
[10] Car Pusic and Diana, "Neural network-based model for predicting preliminary construction cost as part of cost predicting system," Advances in Civil Engineering, vol. 20, no. 20, pp.1-13, 2020.
[11] Swei, Omar, Jeremy Gregory, and Randolph Kirchain, "Construction cost estimation: A parametric approach for better estimates of expected cost and variation," Transportation Research Part B: Methodological, vol. 101, pp.295-305, 2020.
[12] Shahrara, Neda, Tahir Çelik, and Amir H. Gandomi, "Gene expression programming approach to cost estimation formulation for utility projects," Journal of Civil Engineering and Management, vol. 23, no. 1, pp.85-95, 2017.
[13] Peng Duan, "Research on construction project cost prediction method," M.S. thesis, CQU, ChongQing, China, 2008.
[14] Elmousalami and Haytham H, "Artificial intelligence and parametric construction cost estimate modeling: State-of-the-art review," Journal of Construction Engineering and Management, vol. 146, no. 1, 2020.

[15] Jian-Wei Liu, Yuan Liu, and Xiong-Lin Luo, "Advances in deep learning research," Computer Application Research, vol. 31, no. 7, pp.1921-1930, 2014.
[16] JTG/T D70-2010, Design rules for highway tunnels [S].
[17] Warfield and John N, "Binary matrices in system modeling," IEEE Transactions on Systems, Man, and Cybernetics, vol. 5, pp.441-449, 1973.
[18] Zhi-Jun Sun, Lei Xue, and Yang-Ming Xu, "A review of deep learning research," Computer Application Research, vol. 29, no. 8, pp.2806-2810, 2012.
[19] Fei-Niu Yuan, Lin Zhang, and Jin-Ting Shi, "A review of the theory and application of self-coding neural networks," Journal of Computing, vol. 42, no. 1, pp.203-230, 2019.
[20] Olshausen, Bruno A., and David J. Field, "Sparse coding with an overcomplete basis set: A strategy employed by V1?" Vision Research, vol. 37, no. 23, pp.3311-3325, 1997.
[21] Xiao-Ai Dai, Shou-Heng Guo, and An Ren, "Hyperspectral image classification based on stacked sparse self-coder," Journal of University of Electronic Science and Technology, vol. 45, no. 3, pp.382-386, 2016.
[22] JTG/T F60-2009, Technical specifications for highway tunnel construction [S].
[23] Xiao-Long Chen and Li-Guang Wang, "Project cost estimation based on architectural design parametric analysis model," Journal of Tongji University, vol. 37, no. 8, pp.1115-1121, 2009.
[24] Bi-Jun Liu and Yu-Chen Ye, "Construction cost prediction for building projects based on stack noise reduction autoencoder," Journal of Tongji University, vol. 48, no. 6, pp.922-928, 2020.

Jing-Qun Zhou was born in GanSu, China, in 1997. She obtained her Bachelor's degree in Civil Engineering from Lanzhou Jiaotong University in 2019 and has professional experience at Gansu Road and Bridge Construction Group. Presently, she is a graduate student at the School of Transportation at Lanzhou Jiaotong University, where she is mentored by Professor Chang-Xi Ma. Her research interests include the construction and management of civil engineering as well as the organization and planning of transportation engineering.